ORKG

TIB LEIBNIZ INFORMATION CENTRE FOR SCIENCE AND TECHNOLOGY UNIVERSITY LIBRARY

IGB Leibniz-Institut für Gewässerökologie und Binnenfischerei

Beyond keywords

A template for annotating research in ecology using the ORKG platform

Plan for the session

1. Introducing ontologies, knowledge graphs, graph modelling, and the ORKG
2. Motivation
3. Our template model
4. Demo
5. **Practical exercise!**

   break
6. Creating a comparison
7. General feedback on the approach
8. Perspectives for community participation and a joint paper

Why a template for annotating studies?

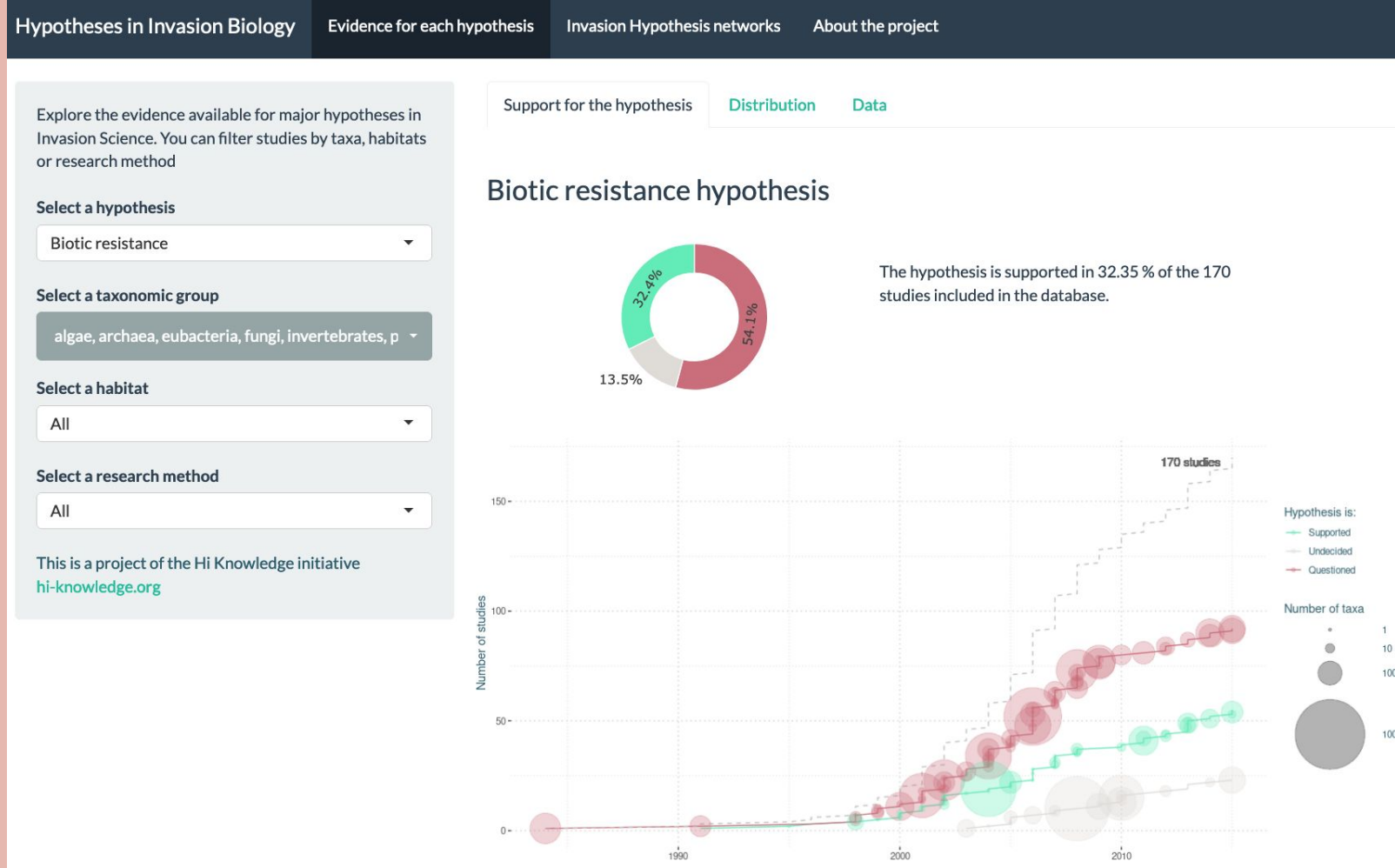
Guiding authors to annotate key information about their study

- **Pre-selected list of common properties**: avoid information gaps
- **Controlled vocabulary**: avoid redundancy, ensure comparability
- **Semantic enrichment**: provide meaning beyond words
- Make it **machine-actionable** thanks to the graph model

Promises automatic & reliable:

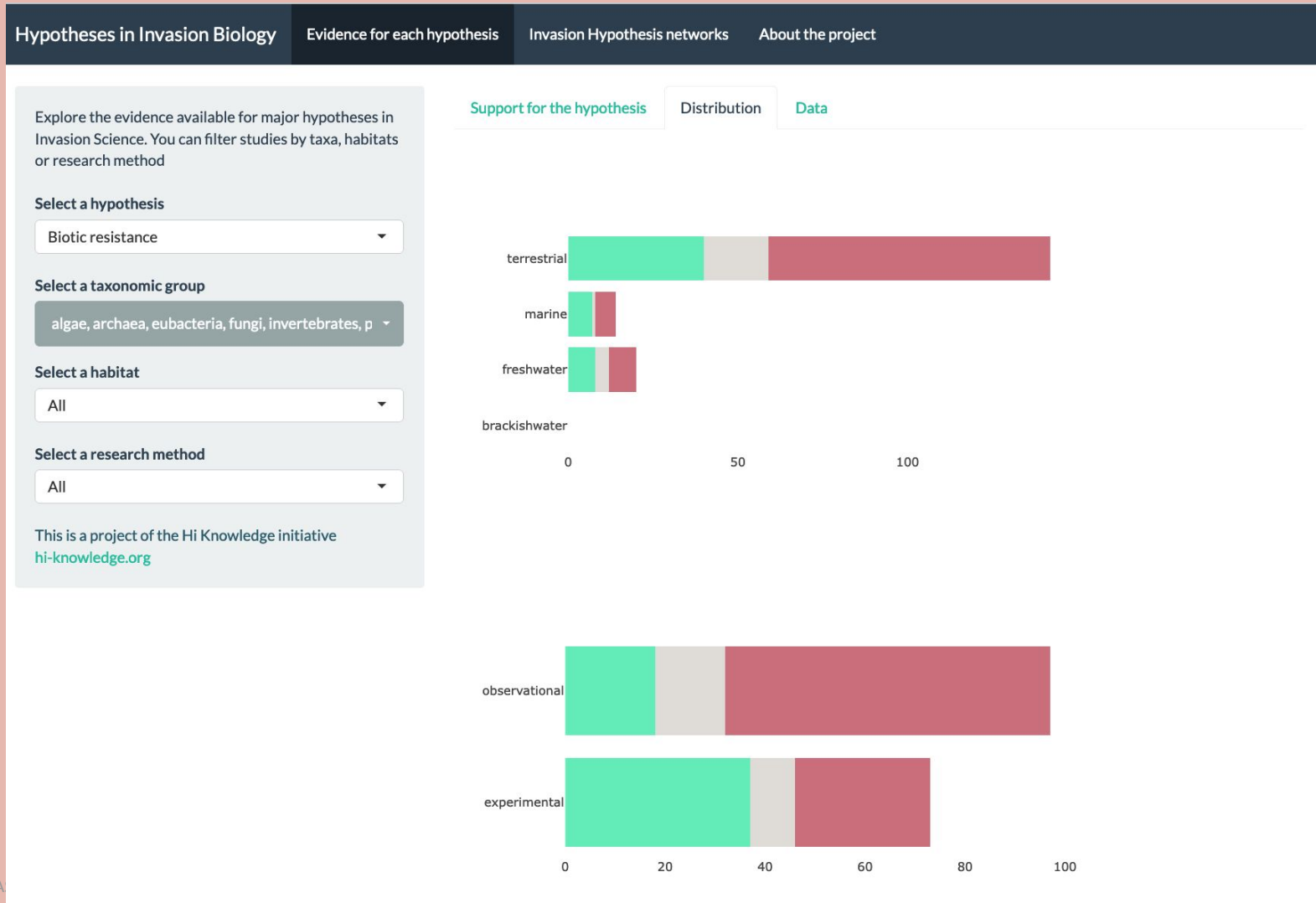
- Classification
- Comparison
- Synthesis or meta-analyses

Automated interactive online synthesis



https://maudbernardverdier.shinyapps.io/Explore_invasion_hypotheses/#

Automated interactive online synthesis



Automated interactive online synthesis

ORKG View Tools About NFDI4DataScience enemy release + Add new

Comparison | 248 contributions

Edit + Add contribution Visualize Actions

Invasion Biology- Enemy release hypothesis

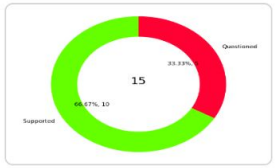
November 2020

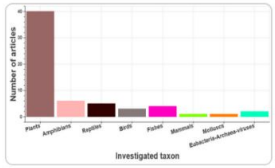
Comparison of studies that are relevant to the Enemy release hypothesis. This hypothesis says: The absence of enemies in the exotic range is a cause of invasion success.

IGB Leibniz-Institut für Gewässerökologie und Binnenfischerei

Invasion Biology

Share





Properties	The invertebrate fauna on broom, Cytisus scoparius, in two native and two exotic habitats	A Comparison of Herbivore Damage on Three Invasive Plants and Their Native Congeners: Implications for the Enemy Release Hypothesis	Can enemy release explain the invasion success of the diploid Leucanthemum vulgare in North America?	Incorporation of an inva plant into a native insec herbivore food web
	Contribution 2 - 2000	*Contribution 1 - 2000*	*Contribution 1 - 2000*	*Contribution 1 - 2000*

https://orkg.org/comparison/R58002/

Identifying key characteristics to annotate

- Most important search criteria
- Main information filtered in a meta-analysis
- Interesting groupings for future synthesis

=> First enKORE/INAS workshop in May 2022

(access full report here)

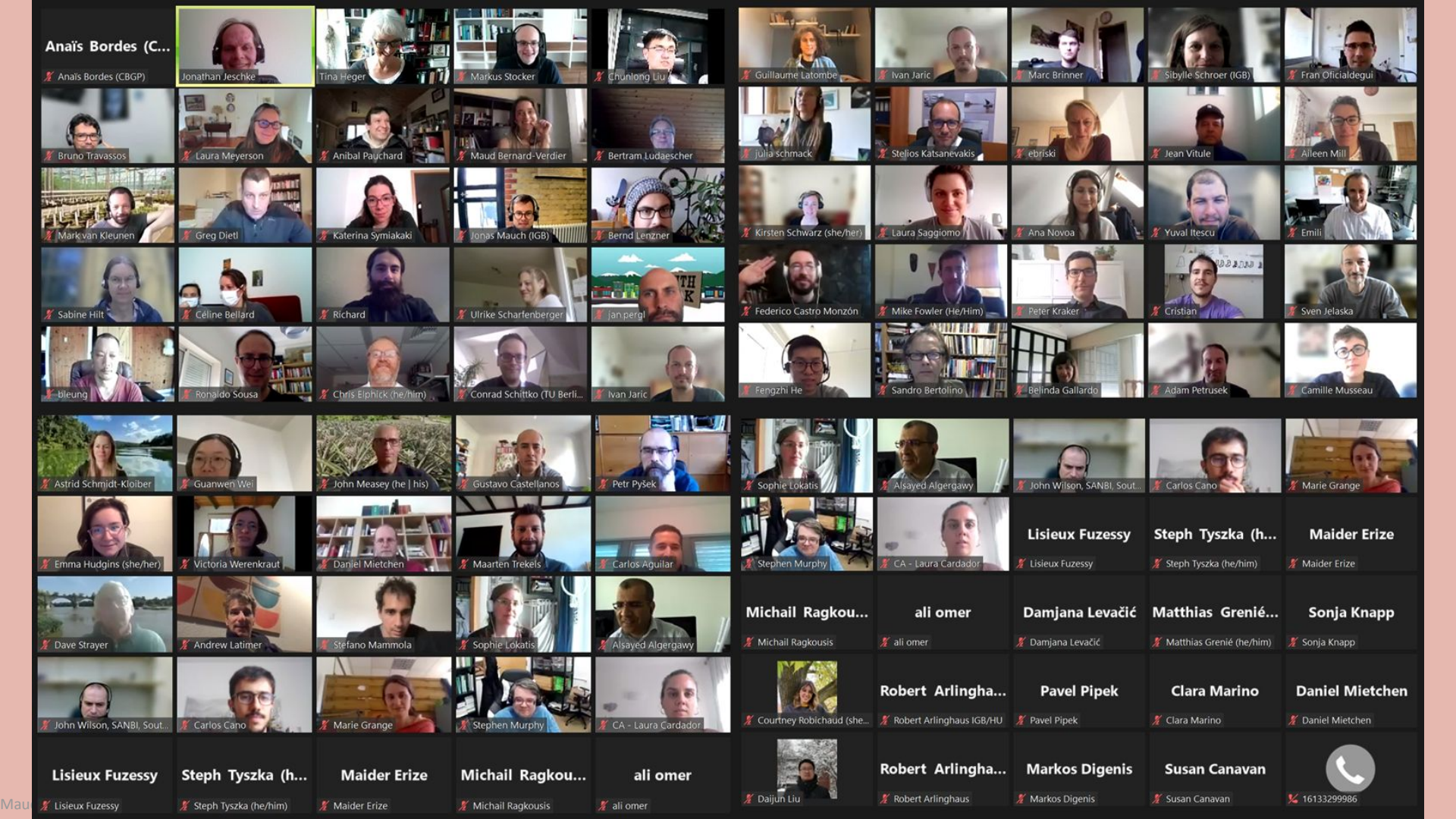

Anaïs Bordes (C...
Anaïs Bordes (CBGP)
Jonathan Jeschke
Tina Heger
Markus Stocker
Chunlong Liu
Guillaume Latombe
Ivan Jaric
Marc Brinner
Sibylle Schroer (IGB)
Fran Oficialdegui
Bruno Travassos
Laura Meyerson
Anibal Pauchard
Maud Bernard-Verdier
Bertram Ludaescher
julia schmack
Stelios Katsanevakis
ebriski
Jean Vitule
Aileen Mill
Mark van Kleunen
Greg Dietl
Katerina Symiakaki
Jonas Mauch (IGB)
Bernd Lenzner
Kirsten Schwarz (she/her)
Laura Saggiomo
Ana Novoa
Yuval Itescu
Emili
Sabine Hilt
Céline Bellard
Richard
Ulrike Scharfenberger
jan pergl
Federico Castro Monzón
Mike Fowler (He/Him)
Peter Kraker
Cristian
Sven Jelaska
bleung
Ronaldo Sousa
Chris Elphick (he/him)
Conrad Schittko (TU Berli...
Ivan Jaric
Fengzhi He
Sandro Bertolino
Belinda Gallardo
Adam Petrusek
Camille Musseau
Astrid Schmidt-Kloiber
Guanwen Wei
John Measey (he | his)
Gustavo Castellanos
Petr Pyšek
Sophie Lokatis
Alsayed Algergawy
John Wilson, SANBI, Sout...
Carlos Cano
Marie Grange
Emma Hudgins (she/her)
Victoria Werenkraut
Daniel Mietchen
Maarten Trekels
Carlos Aguilar
Stephen Murphy
CA - Laura Cardador
Lisieux Fuzessy
Lisieux Fuzessy
Steph Tyszka (h...
Steph Tyszka (he/him)
Maider Erize
Maider Erize
Dave Strayer
Andrew Latimer
Stefano Mammola
Sophie Lokatis
Alsayed Algergawy
Michail Ragkou...
Michail Ragkousis
ali omer
ali omer
Damjana Levačić
Damjana Levačić
Matthias Grenié...
Matthias Grenié (he/him)
Sonja Knapp
Sonja Knapp
John Wilson, SANBI, Sout...
Carlos Cano
Marie Grange
Stephen Murphy
CA - Laura Cardador
Courtney Robichaud (she...
Robert Arlingha...
Robert Arlinghaus IGB/HU
Pavel Pipek
Pavel Pipek
Clara Marino
Clara Marino
Daniel Mietchen
Daniel Mietchen
Lisieux Fuzessy
Lisieux Fuzessy
Steph Tyszka (h...
Steph Tyszka (he/him)
Maider Erize
Maider Erize
Michail Ragkou...
Michail Ragkousis
ali omer
ali omer
Daijun Liu
Robert Arlingha...
Robert Arlinghaus
Markos Digenis
Markos Digenis
Susan Canavan
Susan Canavan
16133299986

Table 1. Characteristics of papers to filter and organize literature searches

Study characteristics	Description and example	Particular issues?	Why is it useful?	Mentioned in break out groups
Basic bibliographic traits				
Article/Publication identifier and content	Automation requires persistent identifiers, preferably DOI; having some additional bibliographic metadata helps with curation	OK Maps needs access to abstracts and/ or full text (PDF/ HTML/ XML), for which Wikidata needs licensing information	we use the papers as the basic unit of building the invasion biology corpus	Wikidata
Authors	List of names and affiliations Ideally with information about author order (first, middle and last authors)	Having unique identifiers and standardizing names. ORCID would help. Most helpful would be an identifier that allows retrieving more information on the authors as needed (e.g. research field, career status)	Identifying invasion biologists vs. other ecologists to help define the scope Identifying senior scientists/leading scientists = an entry point for early career Author co-authorship networks/citation networks to identify communities of researchers	Wikidata Early career
Journal	Name of journal	ISSN	Journal reputation informs trustworthiness of paper Finding papers within a given scope/research context or field	Early career
Citations	Number of citations	Different counts according to reference (WoK, Google scholar…). Needs to be constantly updated automatically.	Useful for early career to assess relevance	Early career

Study characteristics	Description and example	Particular issues?	Why is it useful?	Mentioned in break out groups
Type of paper	Or Article category: review, meta-analyses, original research, opinion paper, perspective paper, methods paper…	Not always tagged in/by the journals?	Reviews and opinion papers are useful to get acquainted with state of the art and research gaps Filtering out secondary analyses when extracting papers for a meta-analyses	Early career Meta-analyses
Useful characteristics				
Habitat	Habitat, or habitats, investigated in the study. May need to allow for multiple entries relating to habitat of origin, invaded habitat, or in cases where multiple habitats are invaded.	No unique classification or ontology of habitats. Can be defined at different levels of resolution (e.g. "terrestrial" vs "calcareous Mediterranean grasslands") Could include the condition of habitat (degraded, restored, intact…)	Main entry point for scanning the literature for early careers	Wikidata Meta-analyses Early career
Taxonomic group of the invader	Taxonomic group of the invader.	different taxonomic resolutions, from coarse polyphyletic names (e.g. "Trees" or "Fish") to the precise sub-species. Synonyms and different taxonomies	Main entry point of lit search (e.g. for early career researchers) Main interest for managers	Wikidata ORKG Early career Freshwater biodiversity
Location	Location of the study sites Ideally very specific information up to the geographical coordinates	May be multiple and hierarchized Often imprecise in the paper (no geographical coordinates) different GIS reference systems: homogenize by using long_lat?	Entry point for literature search (managers,..) Scoping for meta-analysis	Meta-analyses Freshwater biodiversity
Data sets	Direct link to download open data supporting the paper, both underlying data and results. Should include description of datasets (metadata)	Not yet the norm but quickly changing When provided, datasets are not always complete/well described	Useful for meta-analyses aiming to re-analyse the data	Meta-analyses Early career Wikidata
Spatial scale	Spatial scale of the main phenomenon studied (e.g. invasion impact on a local community/landscape/global)	Sometimes difficult to define can be multiple and hierarchized Need to differentiate between grain and extent	scoping meta-analyses	Meta-analyses Early career
Temporal scale	Length of the dataset used in the study, or temporal scale of the phenomenon under study	may be multiple Both grain and extent Needs a unit	scoping meta-analyses	Meta-analyses
Approach	General methodological approach: experimental, observational, theoretical/mathematical	Should be very generic, but perhaps allow for hierarchical structure with a subcategory (e.g. experimental>greenhouse, or theoretical>simulations).	scoping meta-analyses Reviewing methodologies	Early career
Describing methods and results				
Methods	Methods used in the paper, which could be field (e.g. Braun-Blanquet vegetation cover assessment, Capture-Recapture), lab (e.g. AFLP sequencing) or statistical (e.g. Random Forests, GLMMs, etc.) methods.	standardisation of terms is not obvious, but will be necessary to be useful balance between generality and specify may be multiple and hierarchized by types A more precise description would provide a structure for reporting	scoping meta-analyses finding information about a methodology when early career	Meta-analyses Early career

Study characteristics	Description and example	Particular issues?	Why is it useful?	Mentioned in break out groups
		results=> metrics, units, and statistics		
Sample size	Number of replicated units of study	May be of different kinds (plots, species, individuals) Usually multiple and hierarchical: need an explicit data model to describe the study design	scoping and analysing meta-analyses	Meta-analyses
Sampling unit	Object or scale at which measurements are replicated, e.g. organ, individual, colony, population, plot, river, etc.	Very different kind of units are possible Some might need to be described by a unit of measurement (e.g. 5 m x 5m plots vs. 1Ha forest plots).	Scoping meta-analyses	Meta-analyses
Metrics	Which measurements were made or indices calculated?	Very heterogeneous Should be associated to a unit and a method	Meta-analyses Methodological reviews	Meta-analyses
Advanced concepts				
Hypothesis	Research hypothesis, as in Hi-Knowledge	There is no standardized way yet for describing hypotheses	Scoping meta-analyses Reviewing current support for the hypothesis Navigating hypotheses and finding knowledge gaps	Wikidata
Research question/Problem	Overall research question or problem that the study is trying to address. It is related to the research context (see below)	Genericity vs. Specificity Could be designed as a hierarchy	Entry point for managers Entry point for early career and anyone with new topic Navigating and finding quickly current answers to this problem in the literature (when crossed with other filters)	Wikidata Freshwater biodiversity
Index of controversy	An index describing whether a paper has encountered controversy	To be defined Cf. Scholia has a preliminary version of this	Evaluating the trustworthiness and importance of a piece of a theory or a result, especially for early career	Early career Wikidata

Selecting a subset of characteristics to model

Bibliometric data is already mostly well covered

Focus on the **“most important” and “easiest” attributes**

=> Common to all ecology and evolution studies

Add a few more **advanced concepts specific to invasion biology**

=> Theme/Research question/Hypotheses

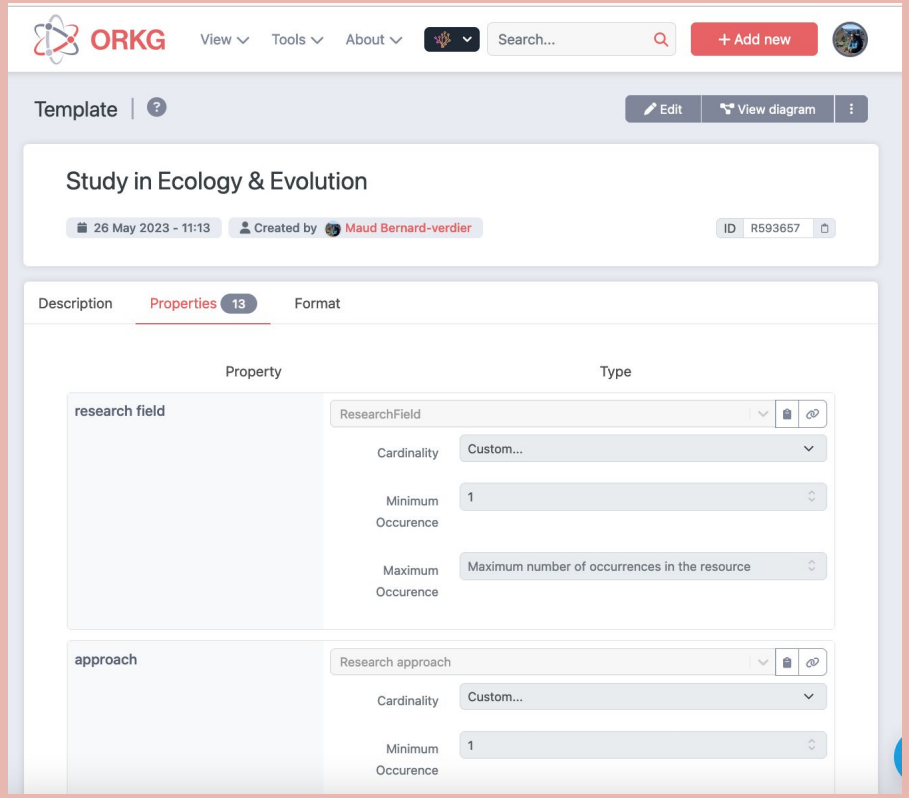
Providing *competency questions* to guide modelling

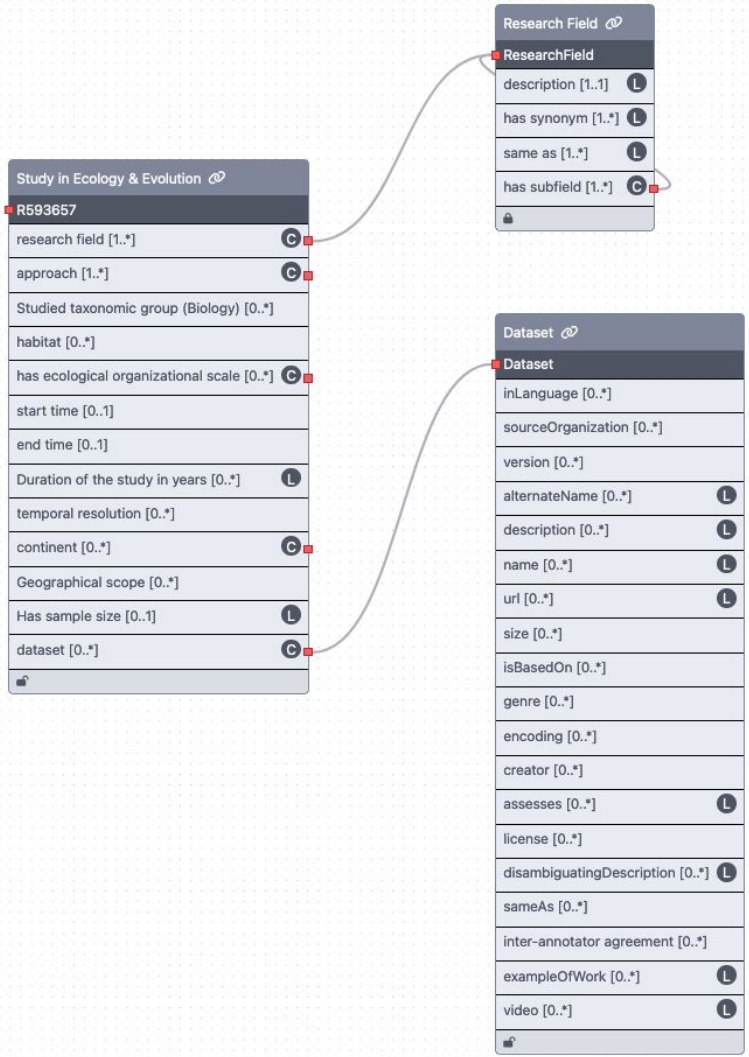
Subject of the template	**Item in the graph model?**	**Statement examples**	**Useful characteristic it provides**
study system	*Research input*	The study system lived in terrestrial habitats	*habitat*
		The study system included plants	*Taxonomic group*
		The study system is an ecological community	*Ecological scale*
		The study system is a plant community	*This provides info both about taxa and ecological scale*
study method	*Research method*	The study method is empirical	*Approach*
		The study method is experimental	*Approach*
		The study method has three steps: 1. Step one of the study method is transplanting 2. Step two of the study method is measuring abundances (the vegetation) 3. Step three of the study method is measuring reproductive fitness	*Optional detail of methods, with sequence of methods*
study	Research activity	The study was repeated every year	Temporal scale : granularity
study	Research activity	The study was carried out between 2002 and 2012	Temporal scale : extent

Initial graph model by Lars

(see Miro board)

The ORKG template for studies in ecology and evolution





Examples and demo

(go to my ORKG profile)

Practical exercise:
annotate your own paper in ORKG

Plan for the exercise

Individual work (call one of us for troubleshooting!):

1. **Annotate** one of your own papers in ORKG using the main template (20 mins)
2. Try annotating with one of the **other templates** (10 min)
3. **Add your links** to the paper webpage in ORKG in the google doc for this session

Group work:

1. General round of feedback
2. Assemble contributions in a comparison
3. Final discussion:
    - What worked?
    - What is not working/weird?
    - What is missing?

How to annotate your paper in ORKG

1. Get started:

- Go to orkg.org
- Create a free account

2. Add a paper in ORKG:

- Choose your favorite paper and get its DOI/url
- On the ORKG main home page click on +Add new > Paper
- Enter DOI > look up
- Choose research field life science> Ecology and evolutionary biology
- Your paper is now added and save it by clicking "finish"
- Click on Edit
- Check the biblio data is correct in Edit data (below paper title)
- Click on X stop editing

You are now ready to annotate your paper!

3. Annotate your paper with the main template:

- Click on Edit
- Select Template
- Choose template: "Study in ecology and evolution"
- Fill out the fields as best you can!
  - Hover over each property to read the description
  - choose pre-existing options whenever possible
  - Choose recommended Wikidata entries (marked by a star) whenever possible
- Save by clicking on "finish"

4. Try out further templates:

- "Invasion biology research questions": Select theme, research question and hyp from a drop-down menu
- Ecological study system
- Ecological study design

Assembling a comparison

Go to existing comparison:

https://orkg.org/comparison/R595599/

Feedback & Perspectives

Feedback on the ORKG template

- What worked?
- What did not work/was strange and needs to be fixed?
- What is unnecessary?
- What is missing?
- How should the template be organised/behave?
    - Multiple sub-templates vs. one large overall template?
    - Stay at a very general level vs. pre-select detailed properties

What future for such manual annotations?

- How far are we from such manual annotations with existing tools?
- How to get the community to realistically use such a template?
- What would it take for you to fill out such a template for each new publication?
- What other possibilities do you know/imagine for improving semantic annotation of papers?

Joint publication idea

A call to the community:

Beyond keywords: semantic modelling for better annotation of studies in invasion ecology

- Argue for the importance of better annotations of ecological research
- Propose our template
- Present our method for building it
- Illustrate with a comparison/interactive analyses platform (R shiny)
- End on a call to the community & publishers ?